

# References

- <sup>1</sup> Etter JE, Perneger TV, Ronchi A. Collecting saliva samples by mail. *Am J Epidemiol* 1998;**147**:141–46.
- <sup>2</sup> Sexton M, Nowicki P, Hebel JR. Verification of smoking status by thiocyanate in unrefrigerated, mailed saliva samples. *Prev Med* 1986; **15**:28–34.
- <sup>3</sup> Holden WE, Bartos F, Theime T, Green TR, Fitchen JH. Theophylline in oral mucosal transudate. A practical method for monitoring outpatient therapy. *Am Rev Respir Dis* 1993;**147**:739–43.
- <sup>4</sup> Clements AD, Parker CR. The relationship between salivary cortisol concentrations in frozen versus mailed samples. *Psychoneuroendocrinology* 1998;**23**:613–16.
- <sup>5</sup> Thompson SK. *Sampling*. New York: John Wiley & Sons, 1992.
- <sup>6</sup> Kelly A, Sinclair H. Deprivation and health: identifying the black spots. *J Health Gain* 1997;**1**:13–14.
- <sup>7</sup> *Epi Info, Version 6*. Atlanta, GA: Centers for Disease Control, 1994.
- <sup>8</sup> National Immunisation Committee. *Immunisation Guidelines for Ireland. 2nd Edn*. Dublin: Royal College of Physicians, 1999.
- <sup>9</sup> Asch DA, Jedrzewski MK, Christakis NA. Response rates to mail surveys published in medical journals. *J Clin Epidemiol* 1997;**50**: 1129–36.
- <sup>10</sup> Kish L. Methods for design effects. *J Offic Stat* 1995;**11**:55–77.
- <sup>11</sup> Central Statistics Office. *Census 1996: Principal Demographic Results*. Dublin: Stationery Office, 1996.
- <sup>12</sup> Friel S, Nic Gabhainn S, Kelleher C. *The National Health and Lifestyle Surveys*. Galway: National University of Ireland, Centre for Health Promotion Studies, 1999.
- <sup>13</sup> Fox RJ, Crask MR, Jonghoon K. Mail survey response rate, a meta-analysis of selected techniques for inducing response. *Pub Opin Quart* 1988;**52**:467–91.
- <sup>14</sup> Stata Corporation. *Stata Statistical Software: Release 5.0*. College Station, TX: Stata Corporation, 1997.
- <sup>15</sup> Eltinge JL, Scribney WM. Some basic concepts for design-based analysis of complex survey data. *Stata Tech Bull* 1996;**31**:3–6.
- <sup>16</sup> Allwright S, Barry J, Bradley F, Long J, Thornton L. *Hepatitis B, C and HIV in Irish Prisoners*. Dublin: Government Publications Office, 1999.
- <sup>17</sup> Smyth BP, Keenan E, O'Connor JJ. Blood borne viral infection in Irish injecting drug users. *Addiction* 1998;**93**:1649–56.
- <sup>18</sup> Devlin JB, Mulcahy M, Corcoran R. Hepatitis B in the non-residential mentally handicapped population. *J Intellect Disabil Res* 1993;**37**: 553–60.
- <sup>19</sup> Starke K, Warnecke C, Brinkmann V. Sensitivity of HIV antibody detection in saliva. *Med Microbiol Immunol* 1993;**182**:147–51.
- <sup>20</sup> Richards AL, Perrault JG, Caringal LT *et al*. A non-invasive assessment of hepatitis B virus carrier status using saliva samples. *Southeast Asian J Trop Med Public Health* 1996;**27**:80–84.

# Commentary: Extending the boundaries of data collection by mail

Thomas V Perneger<sup>a,b</sup> and Jean-François Etter<sup>a</sup>

In this issue of the *International Journal of Epidemiology*, O'Connell *et al.* report on a population-based study estimating the prevalence of hepatitis B virus seropositivity in Ireland, in which saliva samples were collected by the respondents themselves, and mailed back to the researchers.<sup>1</sup> Although the cost of the study is not reported, we may guess that it was cheap: respondents collected the samples for free, and sample delivery was the cost of a regular mail stamp. What is remarkable is not that this study was done, but that such mail-based studies are not done more often.

Using the study participant as the primary data collector and the mail as route of delivery is routine for questionnaire surveys, but not for gathering other types of data. Possible uses (some tested, some not) are numerous. Saliva samples can be used not only for serological testing, but also for detection of various other biomarkers, such as cotinine.<sup>2</sup> If oral mucosa is scraped

in order to collect epithelial cells, the material allows for genetic testing,<sup>3</sup> thus opening the field of population-based genetic studies. Other bodily fluids could be presumably collected in this way, including nasal secretions,<sup>4</sup> urine,<sup>5</sup> stools,<sup>6</sup> and perhaps even blood samples, at least in select subgroups such as patients with diabetes. Collecting nail clippings (e.g. for arsenic determination<sup>7</sup>) or hair (e.g. for substance abuse measurements<sup>8</sup>) might be even easier.

Furthermore, respondents may be asked to provide health-related data based on self-inspection (e.g. the number of missing or filled teeth checked in the mirror<sup>9</sup>) or simple self-measurement (e.g. peak-flow meter readings, skinfold thickness) if the measurement instrument can be supplied. Respondents could also provide data on environmental exposures: household dust (perhaps collected in freshly changed vacuum bags, to test for allergens), samples of tap water (to measure magnesium content, or bacteriological purity), chips of housepaint (in search of lead), and perhaps even indoor air samples. Participants might also provide remnants or wrappings of consumption products, such as cigarette butts, empty cigarette packs, or medication packaging, as a way of validating self-report or assessing the precise

<sup>a</sup> Institute of Social and Preventive Medicine, University of Geneva, and <sup>b</sup> Quality of Care Unit, Geneva University Hospitals, Geneva, Switzerland.

content of such products. Given the low cost of single use photographic cameras, participants could be asked to take pictures of their open refrigerators (to study food content<sup>10</sup>), medicine cabinets (to assess self-medication patterns), housing arrangements (to check for hazards increasing the risk of falls), or even of themselves (to assess results of skin treatments).

Such studies transfer onto the study participant the burden of collecting primary data, but also the responsibility for data quality. This reliance on untrained data collectors may cause concern. Much like instructions that accompany cheap Swedish self-assembly furniture, instructions for conducting sampling procedures and measurements must be clear and foolproof, and the instruments cheap and reliable. The study by O'Connell suggests that saliva collection is feasible, as more than 98% of samples were suitable for testing. The question of feasibility remains open for other possible uses of autonomous data collection. Furthermore feasibility does not guarantee quality. To what extent autonomous data collection introduces information bias is unclear. Other problems might arise from inadvertently or wilfully incorrect data collection procedures being used by the distracted or the annoyed.

However, the main limitation of such mail-based do-it-yourself data collection methods is selection bias. Firstly, in most postal surveys, a small percentage of responders are not those for whom the questionnaire was intended. Whether this percentage would increase when the data collection is more complex than checking all boxes that apply remains to be seen. Much more important is global non-response. In a survey of smokers, the request of a saliva sample from a random subsample of participants substantially reduced the initial response rate.<sup>2</sup> The reduction in the response rate may be greater for more challenging data collection procedures, or when the procedure is perceived as potentially threatening for the respondents' privacy. For instance, even if O'Connell *et al.* clearly stated the purpose of their survey, some people might have feared that testing for other viruses, such as HIV, would be performed. Guaranteeing the confidentiality and anonymity of sensitive data in such studies is crucial.

Even if data collection is simple and non-threatening, mail surveys may produce insufficient response rates. The study by O'Connell reached 60%, which is commendable, given that saliva samples were requested. This response rate may have been bolstered by the offer of a free lottery prize, and more imaginative research is needed to identify incentives that work best in general population settings. Nevertheless, we may wonder how meaningful a prevalence rate estimate of hepatitis B virus antibodies of 0.5% is when based on only five positive tests. How likely is it that the 40% of non-respondents would have the same risk exposure patterns, knowing that exposure to hepatitis B is likely more frequent among drug users, Asian immigrants, or other strata of society who are less likely than others to participate in such a survey? More subtly, the reason

for non-participation may be important;<sup>11</sup> people who decline participation because they resent providing a saliva sample may differ from those who hate filling questionnaires or those who lack the time (incidentally, the possibility that incentives aimed at increasing the response rate may in themselves cause information or selection bias in mail surveys remains largely unexplored). The consequences of partial participation may be less serious if the study under consideration is analytical rather than descriptive; in other words, associations between variables may be less sensitive to selection bias than prevalence estimates or other descriptive statistics.<sup>12</sup>

Even in this age of the internet, mail-based studies remain appealing: almost everyone has a postal address and most people have only one, sampling frames are fairly up to date, most people read their mail in timely fashion, communication costs are reasonable, and, as O'Connell *et al.* demonstrate, more than written information can be sent over that old-fashioned communication network.

## References

- O'Connell T, Thornton L, O'Flanagan D *et al.* Oral fluid collection by post for viral antibody testing. *Int J Epidemiol* 2001;**30**:298–301.
- Etter JE, Perneger TV, Ronchi A. Collecting saliva samples by mail. *Am J Epidemiol* 1998;**147**:141–46.
- Walker AH, Najarian D, White DL, Jaffe JE, Kanetsky PA, Rebbeck TR. Collection of genomic DNA by buccal swabs for polymerase chain reaction-based biomarker assays. *Environ Health Perspect* 1999;**107**:517–20.
- Kawamoto E, Sawada T, Maruyama T. Evaluation of transport media for *Pasteurella multocida* isolates from rabbit nasal specimens. *J Clin Microbiol* 1997;**35**:1948–51.
- Macleod J, Rowsell R, Horner P *et al.* Postal urine specimens: are they a feasible method for genital chlamydial infection screening? *Br J Gen Pract* 1999;**49**:455–58.
- Rothenbacher D, Inceoglu J, Bode G, Brenner H. Acquisition of *Helicobacter pylori* infection in a high-risk population occurs within the first 2 years of life. *J Pediatr* 2000;**136**:744–48.
- Ovaskainen ML, Virtamo J, Alfthan G *et al.* Toenail selenium as an indicator of selenium intake among middle-aged men in an area with low soil selenium. *Am J Clin Nutr* 1993;**57**:662–65.
- Kuhn L, Kline J, Ng S, Levin B, Susser M. Cocaine use during pregnancy and intrauterine growth retardation: new insights based on maternal hair tests. *Am J Epidemiol* 2000;**152**:112–19.
- Perneger TV, Whelton PK, Klag MJ. Race and end-stage renal disease: socio-economic status and access to health care as mediating factors. *Arch Intern Med* 1995;**155**:1201–08.
- Boumendjel N, Herrmann F, Girod V, Sieber C, Rapin CH. Refrigerator content and hospital admission in old people. *Lancet* 2000;**356**:563.
- Etter JE, Perneger TV. Analysis of non-response bias in a mailed health survey. *J Clin Epidemiol* 1997;**50**:1123–28.
- Etter JE, Perneger TV. Snowball sampling by mail: application to a survey of smokers in the general population. *Int J Epidemiol* 2000;**29**:43–48.